EECS 126 Notes

Agnibho Roy

Spring 2020

The following is a compilation of my lectures notes for the spring rendition of EECS 126 taught by professor Kannan Ramachandran. I only kept up typing up my notes in the first couple lectures, and got a little lazy and wrote the rest by other means. Currently working on transcribing them and should be done in a couple of weeks.

Contents

1 Lecture 1 - 1/21/2020 **3**
- 1.1 Introduction . . . 3
- 1.2 Sample Spaces . . . 3
- 1.3 Basic Probability Concepts . . . 3

2 Lecture 2: 1/23/2020 **5**
- 2.1 Conditional Probability . . . 5
- 2.2 Birthday Paradox . . . 6
- 2.3 Independence . . . 6

3 Lecture 3: 1/28/2020 **7**
- 3.1 Discrete Random Variables . . . 7
- 3.2 Expectation and Variance . . . 7
- 3.3 Popular RVs . . . 7

4 Lecture 4: 1/30/2020 **8**

5 Lecture 5: 2/04/2020 **9**

6 Lecture 6: 2/06/2020 **10**

7 Lecture 7: 2/11/2020 **11**

8 Lecture 8: 2/13/2020 **12**

9 Lecture 9: 2/18/2020 **13**

10 Lecture 10: 2/20/2020 **14**

11 Lecture 11: 2/25/2020 **15**

12 Lecture 12: 2/27/2020 **16**

13 Lecture 13: 3/03/2020 **17**

14 Lecture 14: 3/05/2020 **18**

15 Lecture 15: 3/10/2020 **19**

16 Lecture 16: 3/12/2020 **20**

17 Lecture 17: 3/17/2020 **21**

18 Lecture 18: 3/19/2020 **22**

19 Lecture 19: 3/31/2020 **23**

20 Lecture 20: 4/02/2020 **24**

21 Lecture 21: 4/07/2020 **25**

22 Lecture 22: 4/09/2020 **26**

23 Lecture 23: 4/14/2020 **27**

24 Lecture 24: 4/16/2020 **28**

25 Lecture 25: 4/21/2020 **29**

26 Lecture 26: 4/23/2020 **30**

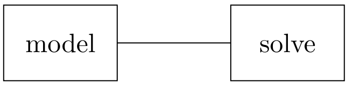
27 Lecture 26: 4/28/2020 **31**

28 Lecture 26: 4/30v/2020 **32**

1 Lecture 1 - 1/21/2020

1.1 Introduction

This class is a class that deals with uncertainty, and has roots in predictions, strategy and decision making, and design. We can set up any problem in this class in the following way:



Here, we use the model to attempt to understand the problem at hand, and then use a set of tools to solve the question using the model that we built. An example of this is answering the question of how many customers we expect to arrive within an hour at the front of a store. We first model the arrivals to the store as a Poisson Process, determining the rate λ of arrivals based on experimental observations. Then we can attempt to solve the question using our model by calculating the expectation of $E[\text{Poisson} \sim \lambda(1)]$

1.2 Sample Spaces

In order to understand probability, we introduce the notion of a **sample space (Ω)**, which is the set of all possible outcomes of an experiment. Two important properties of sample spaces are they are **mutually exclusive** and **collectively exhaustive**, which means that no two samples in the sample space can occur at the same time in one iteration of the experiment, and the sample space sufficiently represents all possible outcomes of the experiment, respectively. Ex. A single toss of two fair coins gives has a sample space $\Omega = \{HH, HT, TH, TT\}$, representing all possible outcomes for the single experiment. Sample spaces can be infinite at times; take for example tossing a coin until the first heads, which has has a countably infinite sample space $\Omega = \{H, TH, TTH \dots\}$ (also shows that all samples need not have the same likelihood of occurring) or the amount of time you have to wait for a bus to come at a bus stop that will come at latest at some time T, which has a uncountably infinite sample space $\Omega = (0, T)$.

An **event** is defined as any allowable subset of Ω that combines. The probability of these events is the sum of the individual probabilities of the elements in the sample space that are in the event. Another way to think about it is that the events are mapped to a probability that is proportional to Ω in its weight. Note that this means that the total possible events is $2^{|\Omega|}$.

Example 1.2.1: Consider the experiment of a single toss of two fair coins. Let the event A = We get at least 1 head. Then:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{P(\{HT, TH, HH\})}{P(\{TT, HT, TH, HH\})} = \frac{3}{4}$$

All of these concepts together form a summarized notion of what a probability space is for an experiment and can essentially be treated as the "cheat sheet" for any question pertaining to the experiment. In general, we obtain the following definition:

Definition 1.2.1: The **probability space** for an experiment can be concisely summarized as (Ω, F, P) where Ω denotes the sample space, F denotes the set of all possible events, which are subsets of Ω, and P denotes the probability of each $\zeta \in F$; an injective mapping of $F \to (0, 1)$

1.3 Basic Probability Concepts

there are some basic rules that all probability spaces should follow, and they are defined by Kolmomgorov as the axioms of probability:

Definition 1.3.1: The **axioms of probability**:

1. $P(\emptyset) = 0$
2. $P(\Omega) = 1$
3. For mutually exclusive events $A_1 \dots A_n$, $P(\bigcup\limits_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

These axioms should be followed like a religion, but there are other fundamental probability facts that are derived from the ones above that should be known, one will find that they are quite useful in problem solving.

Definition 1.3.2: Fundamental Facts of Probability:

1. Complement: $P(A^c) = 1 - P(A)$
2. Union of Events: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. Union Bound: $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$
4. Inclusion/Exclusion Principle: (insert formula)

The most important of these facts is the **Union Bound**, which we can derive by argument with a visual argument using Venn Diagrams. Consider two Venn diagrams representing two separate events, A_1 and A_2. We know that when we take the sum of the two events, we count the middle twice, essentially over counting the intersection by a factor of 2. We can generalize this to n events, where there are a various number of intersections now, namely 2^n possible intersections that could be double counted (can you see why?). Of course, when we have mutual exclusivity of events (intersection is $\emptyset$), then the definition 1.3.2 (3) turns into the equality in definition 1.3.1 (3). The Union Bound is quite useful when it becomes difficult to properly calculate intersections. Here is an example of using the union bound as well as the notion of the considering the complement of an event.

Example 1.3.1: Consider a sphere that has $\frac{1}{10}$th of its surface colored blue, and the rest is colored red. Show that, no matter how the colors are distributed, it is possible to inscribe a cube in the sphere with all of its vertices red. Instead of proving if it is possible for all of the vertices to be red, let us prove that it is impossible that all of the vertices must be blue. Let us create $B_1, B_2, \ldots B_8$ that each represents the events that the *ith* vertex is blue. We compute:

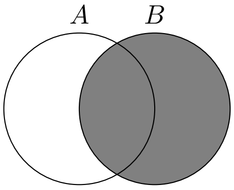
$$P(B_1 \cup B_2 \cup \ldots B_n) \leq \sum_{i=1}^{8} P(B_i) = \frac{8}{10}$$

This is exactly what we wanted to show: that the probability that at least one vertex lies in a blue region is less than 1, because "probability" in this context essentially means if we *randomly* place the cube, there is *at least* a $\frac{2}{10}$ chance that all its vertices lie in a red region. Just rearrange the cube and place it so that this chance happens all the time. If the probability was greater than 1, then its impossible to rearrange the cube in any way.

2 Lecture 2: 1/23/2020

2.1 Conditional Probability

Conditional probability is the probability of an event, but with the knowledge of some "additional information", then we know the. Here is an intuitive example: if I had the knowledge that someone is a really good basketball player, then my probability of them getting. Let us generalize the notion of conditional probability a little more. Consider the two events below:



Let us say that we have the knowledge that B occurs, then we know that the white shaded region is not possible because it is not mutual with B. Therefore, the probability of A occurring now is only the intersection $A \cap B$. We are only halfway done. Moreover, we know that the new "sample space" that we are dealing with is only the samples that occur in B, because all the samples in $A \setminus B$ have no probability now. Thus, we have to *scale* our probability by $P(B)$. This leads to the formula for conditional probability:

Definition 2.4.1: Conditional Probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Conditional probability allows us to express intersections as a chain of conditionals and also allows us to split up a problem into multiple pieces and recombine to the get the final answers. We introduce three equations that relate conditional probabilities:

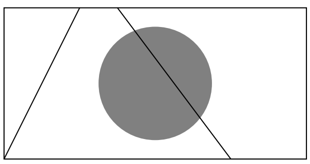
Definition 2.4.2: We can split up the intersection of multiple events $A_1 \dots A_n$ using the **Product Rule**:

$$P(A_1 \cap A_2 \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

If the set of events A_i *partition* the total sample space Ω, meaning that they are all mutually exclusive but span the whole probability space, then we define the **Total Probability Rule** as:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) \dots P(A_n \cap B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

We can prove the **product rule** by induction. Let us start with the base case of $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$, which is just a rearrangement of definition 2.4.1. Let's now consider $P((A_1 \cap A_2) \cap A_3) = P(A_1 \cap A_2)P(A_3|A_2 \cap A_1) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1)$. You can convince yourself that continuing this pattern will give the generalized form in 1.4.2. The **Total Probability Rule**, on the other hand, can be best explained through a visual example:



We can see that from left to right, the rectangle (sample space) is partitioned into events $A_1, A_2, \text{and} A_3$. Moreover, the shaded region is just the intersection of the circle with each of these partitioned, which means that for general n, the shaded region can be represented by $\sum_{i=1}^{n} P(A_i \cap B)$, which can be rearranged using definition 2.4.1 to give the final result in definition 2.4.2 (2)

Finally, Bayes Theorem is a combination of definition 2.4.1 and definition 2.4.2 (2). Let us look at a simple exam of Bayes:

Example 2.4.1: If a person has a disease, then they test positive for the disease with probability 0.95. If they do not have the disease, then they test negative with probability 0.95. Find the probability that the person has the disease, given that the person tests positive. Let A be the probability that

2.2 Birthday Paradox

The Birthday paradox is not really a "paradox" in the sense that once someone does the calculations, the result is quite intuitive. The question asks what is the probability that at least 2 people in a group of n share the same birthday. It turns out that with just 23 people, this probability is $\approx 50\%$. The problem statement is $P(\text{at least 2 people in a group share same birthday}) = 1 - P(\text{no two people share the same birthday})$. Let A represent the left side of the equality and $1 - A^c$ represent the right side. To compute A^c, we know that once the first person takes a birthday, the second person has one less option first because he cannot have the birthday of the first person, the third person has two less options than the first, and so on. Let $k = 365$. This means we have:

$$\begin{aligned}
P(A^c) &= (\frac{k}{k})(\frac{k-1}{k})\dots(\frac{k-(n-1)}{365}) \\
&= 1(1-\frac{1}{k})(1-\frac{2}{k})\dots(1-\frac{n-1}{k}) \\
&= e^{\frac{-1}{k}}e^{\frac{-2}{k}}\dots e^{\frac{-(n-1)}{k}} && (1) \\
&= e^{\frac{-1}{k}(1+2+\dots+n-1)} = e^{\frac{-1}{k}(\frac{n(n-1)}{2})} \approx e^{\frac{-n^2}{2k}} && (2)
\end{aligned}$$

Where in (3) we used the Taylor approximation that for $|x| << 1$, $e^x \approx 1 + x$. Thus, $P(A) = 1 - e^{\frac{-n^2}{2k}}$. Plug in some numbers for n to see how this probability changes.

2.3 Independence

Two events are independent if the occurrence of one provides no information about the occurrence of the other. Mathematically, this means that $P(A|B) = P(A)$ since adding the condition that B occurs does not change the probability of A occurring. Moreover, this means that. A key note is that if A and B are disjoint, then it does not necessarily imply that they are independent. In fact: $P(A \cap B) = 0 \implies P(A) = 0 \text{or} P(B) = 0$, which is not useful, but this proof does show actually that two disjoint events that have non-zero probability are actually *always* dependent.

3 Lecture 3: 1/28/2020

3.1 Discrete Random Variables

3.2 Expectation and Variance

3.3 Popular RVs

4 Lecture 4: 1/30/2020

5 Lecture 5: 2/04/2020

6 Lecture 6: 2/06/2020

7 Lecture 7: 2/11/2020

8 Lecture 8: 2/13/2020

9 Lecture 9: 2/18/2020

10 Lecture 10: 2/20/2020

11 Lecture 11: 2/25/2020

12 Lecture 12: 2/27/2020

13 Lecture 13: 3/03/2020

14 Lecture 14: 3/05/2020

15 Lecture 15: 3/10/2020

16 Lecture 16: 3/12/2020

17 Lecture 17: 3/17/2020

18 Lecture 18: 3/19/2020

19 Lecture 19: 3/31/2020

20 Lecture 20: 4/02/2020

21 Lecture 21: 4/07/2020

22 Lecture 22: 4/09/2020

23 Lecture 23: 4/14/2020

24 Lecture 24: 4/16/2020

25 Lecture 25: 4/21/2020

26 Lecture 26: 4/23/2020

27 Lecture 26: 4/28/2020

28 Lecture 26: 4/30v/2020